

Pesquisa interdisciplinar desenvolve programa que lê texto em voz alta e sem sotaque

O computador que fala português

PAULO CÉSAR NASCIMENTO
pcnpres@uol.com.br

Um programa de computador capaz de ler em voz alta qualquer texto escrito em português, sem o sotaque inglês característico dos sistemas produzidos fora do país, foi desenvolvido em conjunto por pesquisadores das áreas de lingüística e de engenharia elétrica da Unicamp.

Batizado com o nome de *Aiuruetê* (“papagaio verdadeiro” na língua tupi), o software permitiu que ao longo de seu desenvolvimento fosse também gerada mão-de-obra qualificada para a pesquisa brasileira em ciência e tecnologia de fala – uma área em que o país ainda depende de

sistemas importados.

Iniciado em 1991, a princípio como um estudo de descrição fonético-acústica da língua, no âmbito do Laboratório de Fonética e Psicolinguística (Lafape) do Instituto de Estudos da Linguagem (IEL), o projeto logo caracterizou-se pela transdisciplinaridade e pelo diálogo entre as ciências humanas e exatas. Foi quando, um ano depois, passou a contar com a participação do Laboratório de Processamento Digital de Fala (LPDF) da Faculdade de Engenharia Elétrica (FEEC).

“Estávamos atuando em áreas similares, porém isoladamente. Foi a interação das duas áreas que viabilizou o projeto”, afirma o professor Fábio Violaro, coordenador do

LPDF.

Ele explica que, embora sua equipe já trabalhasse com síntese de fala, os resultados mostravam-se limitados por causa da falta de conhecimentos lingüísticos. O conversor texto-fala até então concebido pelo LPDF baseava-se exclusivamente no processamento de sinais, apenas com a manipulação eletrônica da onda sonora da fala.

Sutilezas fonéticas – Foram os trabalhos do Lafape, especialmente o desenvolvimento de um conversor ortográfico-fônico (Ortofon), que dotaram o software de regras de transcrição ortográfico-fonética e da correta pronúncia das pala-

avras no processo de síntese.

“O *Aiuruetê* é um projeto de pesquisa básica que gerou resultados tecnológicos e de formação de recursos humanos muito importantes”, ressalta Eleonora Cavalcante Albano, coordenadora do Lafape. “O mais interessante é que isso ocorreu pela integração de professores, alunos de graduação, pós-graduação e bolsistas de iniciação científica de áreas distintas em torno de um objetivo comum.”

O Ortofon (um dos sete módulos de processamento das informações textuais do *Aiuruetê*) é um dos diferenciais do software em relação aos sistemas estrangeiros de síntese de fala em operação no Brasil, enfatiza

Projeto gera mão-de-obra qualificada

Fotos: Antoninho Perri

Fábio.

Por incorporar avançados recursos lingüísticos, ele permite ao programa desenvolvido pela Unicamp respeitar as muitas nuances de foneticidade da língua portuguesa, o que o aproxima de uma fala mais natural e o coloca em posição de vantagem sobre os softwares similares.

“Nem sempre a pronúncia das palavras é determinada pela grafia”, lembra Eleonora, citando como exemplo os casos das palavras grafadas com a letra “x”, que pode ter o som de “ch”, de “s”, de “ks” ou de “z”.

Há outras situações complexas, como as diferenças entre as representações gráficas do texto e a maneira como elas se expressam na fala, caso do verbo piloto (ô) e do substantivo piloto (ô), ambos com idêntica grafia mas pronúncias diferentes. Ou ainda as siglas e abreviaturas (não se lê IEL da mesma forma que UTI), ou as expressões numéricas (37.881.532), que têm leitura diferente da de um número de telefone (3788-1532), por exemplo.

“São sutilezas que o programa consegue captar no processo de transcrição fonética das informações textuais”, salienta a coordenadora do Lafape.

Quebra-cabeça sonoro – Nos sistemas tradicionais de voz sintética, como os utilizados para o fornecimento de informações bancárias por telefone, as sentenças são organizadas e reproduzidas a partir de um banco de formado por palavras, o que acaba por limitar o vocabulário dessas máquinas.

No *Aiuruetê* é diferente: a síntese se dá pela concatenação de polifones (trechos sonoros com dois ou mais fonemas) armazenados em um dicionário sonoro com aproximadamente 2.500 diferentes fragmentos de sons extraídos de gravações, como se fossem as sílabas desmembradas das palavras.

Ou seja, para sintetizar uma frase, o software, a exemplo de alguém que montasse um quebra-cabeça sonoro, junta de forma harmônica, com entonação e ritmo, todos os elementos fônicos das palavras que a compõem.

“Esse recurso é quase uma simulação da produção da fala humana e o que assegura ao sistema flexibilidade suficiente para reproduzir textos de qualquer tamanho em língua portuguesa. Basta montar foneticamente quantas palavras sejam necessárias”, esclarece Plínio Almeida Barbosa, engenheiro eletrônico e lingüista, e que ajudou a desenvolver o *Aiuruetê* durante o seu pós-doutorado.

O programa roda em qualquer computador com o sistema operacional Windows e foi desenvolvido com financiamento de aproximadamente R\$ 70 mil da Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo).



Em busca da interação



Violaro, Eleonora e Barbosa : projeto promove diálogo entre as ciências humanas e exatas

O advento da era digital tornou a síntese de fala um dos processos mais importantes na comunicação homem-máquina. Por essa razão, aprimoramentos se sucedem nos esforços científicos para dotar o computador de uma fala humana natural. Os criadores do *Aiuruetê*, por exemplo, querem que seja capaz de assimilar os diferentes dialetos brasileiros. Mas só falar já não basta; é necessário também que os sistemas automatizados reconheçam a voz humana e possam interagir com os usuários.

Pesquisas nesse sentido estão sendo desenvolvidas no Instituto Nacional de Telecomunicações (Inatel), com a colaboração do LPDF, para organização de uma inédita base de fala pública nacional, a partir de gravações dos diferentes ritmos do fa-

lar brasileiro. Os estudos contribuirão para aprofundar o conhecimento dos aspectos fônicos do português falado no Brasil e servirão para a elaboração de softwares de reconhecimento de voz.

“A automatização é cada vez mais intensa e, em breve, as pessoas conversarão com máquinas capazes de entendê-las, e não mais com telefonistas, quando precisarem obter informações”, observa o coordenador do LPDF.

No futuro, também não será surpresa se os computadores puderem reconhecer e compreender as emoções humanas não só por meio da voz, mas porque poderão “ver” a movimentação facial e labial de seus interlocutores.

“Basta que imagens faciais sejam captadas por uma câmera e interpre-

tadas por um software, como faz HAL, o super-computador de 2001 – *Uma Odisséia no Espaço*”, ilustra Plínio, em uma alusão ao cérebro-eletrônico da enigmática obra-prima de Stanley Kubrick. No filme, após fazer a leitura labial da conversa sigilosa de dois astronautas a bordo de uma nave espacial, HAL descobre um plano para desativá-lo e decide matar os autores da trama.

Para quem pensava que computadores assim só caberiam na imaginação de escritores e roteiristas de cinema, avanços como os proporcionados pela equipe da Unicamp mostram que a aproximação entre ciência e ficção ocorre a passos cada vez mais largos. No caso do *Aiuruetê*, em alto e bom som também.